Data-Driven Models for Pluvial Flood Forecasting in Brussels

La prévision des inondations pluviales à Bruxelles via des modèles guidés par les données.

Solomon Seyoum[1], Patrick Willems[1] and Boud Verbeiren[1,2]

[1] Vrije Universiteit Brussel, Department of Hydrology and Hydraulic Engineering, Brussels, Belgium
[2] Brussels Company for Water Management (SBGE/BMWB), Direction Exploitation, Brussels, Belgium

RÉSUMÉ

Les inondations font partie des catastrophes naturelles les plus récurrentes et les plus coûteuses. Dans les zones urbaines, les inondations peuvent avoir de graves conséquences, de par la forte densité et la présence d'infrastructures vitales. Lors de précipitations extrêmes, le ruissellement sur des surfaces imperméables est majoriatirement responsable du dépassement de la capacité de drainage du système d'égouttage, ce qui aboutit in fine à des inondations en contexte urbain. De par la limitation des données et du fait de la structure des modèles appliqués à des bassins spécifiques de modèles physiques et conceptuels détaillés, ces derniers sont souvent limités dans leur valeur opérationnelle. Par ailleurs, l'utilisation des techniques d'apprentissage automatique pour la modélisation prédictive a récemment augmenté. Dans cet article, nous décrivons le développement et l'utilisation de modèles basés sur les données pour prévoir les débits de pointe dans les canaux de drainage de Bruxelles, en Belgique, en tant qu'indicateurs des inondations pluviales dans le contexte du projet FloodCitiSense. Nous avons recueilli pendant plusieurs années des données sur les précipitations et les écoulements de près de 13 stations de mesure de précipitation et de 13 stations de mesure du débit à Bruxelles. Nous avons ainsi créé des modèles fondés sur des données pour prévoir les débits de pointe dans les canaux de drainage. L'utilisation de modèles basés sur les données pour anticiper les inondations pluviales à partir des données disponibles sur les débits et les précipitations s'est montrée prometteuse.

ABSTRACT

Floods are among the most recurrent and costly natural disasters. In urban areas impact of flooding can be severe as they are often densely populated and contain vital infrastructure. In case of extreme rainfall, runoff from sealed surfaces is the dominating mechanism which lead to exceedance of the system's drainage capacity, ultimately resulting in urban pluvial flooding. Data limitation and model structure applied to specific catchments of detailed physically based and conceptual models often limit their operational value. On the other hand, the use of machine learning techniques for predictive modelling has recently increased. In this paper we describe the development and use of data-driven models to forecast peak flows in drainage channels of Brussels, Belgium as a proxy for pluvial flooding in the context of FloodCitiSense project. We collected rainfall and runoff data from about 13 rainfall and 13 flow gauging stations in Brussels for several years and build data-driven models for forecasting peak flows in drainage channels. The use of data-driven models to forecaster pluvial flooding from available flow and rainfall data has shown promise.

KEYWORDS

Data driven models, FloodCitiSense, Flood Early Warning System, Urban pluvial flooding

1. INTRODUCTION

Floods are among the most recurrent and costly natural disasters in terms of extensive economic damage and unprecedented loss of human life. This is apparent in that flooding events are reported more frequently in the media in many places in the world. Particularly in urban areas the impact of flooding can be severe because the areas affected are often densely populated and contain vital infrastructure. The frequency of urban flooding is expected to increase due to increasing urbanization, ageing drainage infrastructure and recognised climate change due to anthropogenic activities. Urbanization causes previously permeable ground to be more impermeable due to developments producing increased sealing of the surface with concrete and other construction materials with as result a dramatic increase of the amount of rainwater running off the surface into drains and sewers. A common feature of sewerage and drainage networks in many cities around the world is that they are old, and their condition is unknown. The capacity to drain the surface water is limited and potentially decreases as result of malfunctioning of the system. In addition, in a considerable number of flood-prone urban areas wetter winters and heavier summer showers are expected to put further pressure on our urban drainage networks as result of climate change (Seyoum, 2013; Alam and Rabbani, 2007; Douglas et al., 2008; Grum et al., 2006; Parliamentary Office of Science and Technology, 2007).

In case of extreme rainfall, fast and abundant runoff from sealed surfaces is the dominating mechanism which can quickly lead to exceedance of the system's drainage capacity, ultimately resulting in urban pluvial flooding. Pluvial flooding may be directly due to overland flow from land saturated by heavy rain. This is more common during long periods of rainfall in winter months, though it also occurs in urban areas during intense summer rainfall. Pluvial flooding can also occur where there is adequate drainage channel capacity, but flow cannot enter the channel at the necessary rate. A good example of this is highway flooding caused by a lack of gully capacity.

Due to the fast onset and localised nature of pluvial flooding, occurring at small temporal and spatial scales, high resolution models and data are needed (Jacobsen, 2011; Bruni et al. 2014; Ochoa-Rodriguez et al. 2015, Boud et al 2018). This also demands a fast simulation of flood forecasts.

The FloodCitiSense project proposes an interactive and cooperative framework consisting of citizens, local authorities, research units and industrial partners aiming at improving the monitoring and management of urban pluvial flooding (Boud et al 2018). Modelling of pluvial flooding is in one of the components to develop the flood early warning framework. This paper focuses on modelling of pluvial flooding for Brussels, Belgium.

Several different types of models have been developed to simulate the complex hydrological process of transforming rainfall to runoff in a catchment. These models can be broadly categorized as empirical, conceptual and physically based on how the complex hydrological processes are represented in the model. These models are extensively described in literature (Clarke, 1973) and more recently in Eldho and Kulkarni (2017) and Devi et al. (2015). Physically based and to some extent conceptual models describe the physical processes of a catchment in a detailed way. Data limitation and the model structure applied to specific catchments, however, often limit their operational value. On the other hand, the use of machine learning techniques for predictive modelling has recently increased (Bontempi et al., 2012). Data-driven models such as artificial network (ANN) and classification models such as Random Forest (RF) capture the complex dynamic of a watershed without requiring large and/or distributed hydrological data sets. In this paper we describe the development and use of data-driven models to forecast peak flows in drainage channels of Brussels as a proxy for pluvial flooding.

2. MATERIALS AND METHODS

The main outcome of FloodCitiSense project will be an urban pluvial flood early warning service for, but also by citizens and city authorities, built upon the state-of-the-art knowledge and methodologies. As part of the urban flood forecasting for Brussels, we aim to develop a data-driven modelling approach at sub-catchment scale.

Determination of the size of the forecast horizon is critical in timeseries prediction (Bontempi et al., 2012). The Predictive ability decreases as the forecasting horizon increases. Model effectiveness relies on parameters tuning, input data length and resolution.

To build data-driven models we need large amount of input and output data for training and testing of the models. Historical rainfall and runoff data are available for Brussels from the early 2000's from Flowbru.be, maintained by Société Bruxelloise de Gestion de l'Eau. Users can download these data in

5-minute resolution up to near real-time. Therefore, our first step was to collect available data and do some quality control checking. We collected both rainfall and runoff data from about 13 rainfall and 13 flow gauging stations for several years (from 2014 to 2018) in 5-minute resolution. The determination of optimal model input plays a key role in model performance since it provides the basic information about the system (Wang et al, 2009). For a rainfall-runoff forecasting model, information in addition to the current time step is needed for the model to perform adequately. Additional information can be derived from previous time step rainfall and runoff.

There may be more than one rainfall station which are measuring rainfall that may influence the flow in a runoff measuring station. We used correlation coefficients to select five rainfall stations which influence a flow in each flow station. For runoff input we used lag time equivalent to the forecast horizon, in this case two hours. We consider sum of the rainfall values for the lag period. After the relevant rainfall gauging stations and the lag times are identified and input and output time series are prepared, data-driven models (a multilayer perceptron ANN and RF) were developed, trained and tested to forecaster flood causing peak flows in the drainage channels.

The Data driven models use rainfall data of 5 most correlated rainfall stations and the lagged flow data of the flow station to forecaster the current flow as shown in the following equation.

$$Q_t = f\left(Q_{t-lag}, \sum_{j=t-lag}^{j=t} RF_{i,j}\right) for\, i = 1 to 5$$

Where Q_t is the current flow at a flow station, Q_{t-lag} is the lagged flow at the station and $RF_{i,j}$ is the of rainfall values for station i and time j.

3. RESULTS AND DISCCUSSION

For each flow station, MPL-ANN and RF models are being trained and tested. More than 200,000 data point were available for training and testing the models for each flow station. For most of the flow stations the data-driven models perform well with R-squared vales ranging from 0.55 to 0.98 for a 2-hour forecast horizon. Their performance increases for a one-hour forecast horizon. The Random-forest models perform better in training, thought their performance drops significantly in testing compared to MLP-ANN models. Training the RF models are faster as they are classification models. Figure 1 and 2 show one of the best results for MLP-ANN models during training and testing for flow station N01.

4. CONCLUSIONS

The use of data-driven models to forecaster pluvial flooding from available flow and rainfall data has shown promise in absence of hydraulic model to forecast flooding. As in the case with data-driven models, under estimation of peak flow is observed in both MLP-ANN and RF models. Researchers suggest that appropriate data transformation improves the performance of such models (Sudheer, et al., 2003). Our next step is to test different data transformation techniques to improve the prediction of peak flows.

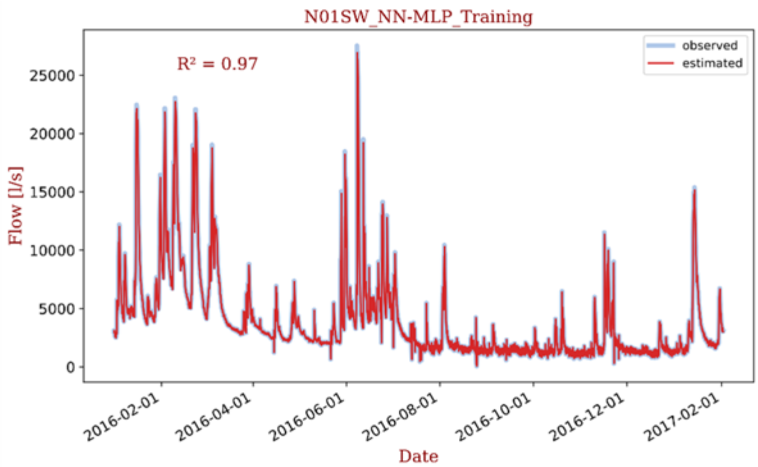

Figure 1- MLP-ANN training result for station N01

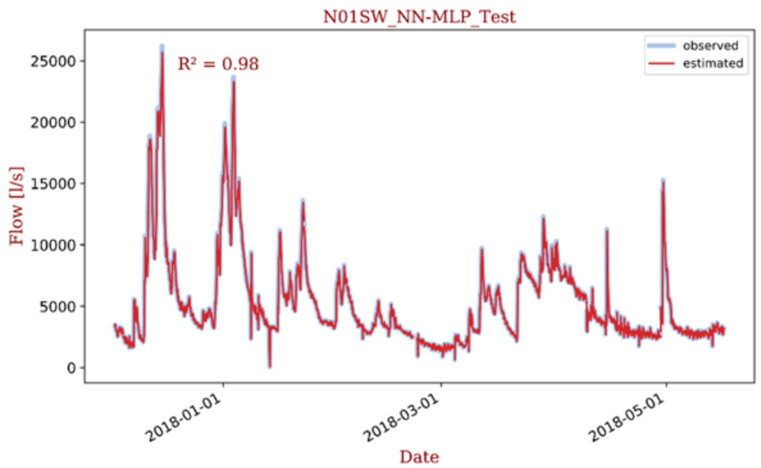

Figure 1- MLP-ANN training result for station N01

The method can be used for other case studies as long as there is enough amount of rainfall-runoff data to train and test the data-driven models. This requirement is one of the limitations of data-driven models. Other limitations include that these methods can only forecast flow at measuring stations and not at another location in the drainage network and cannot be used to forecast depth and spatial extent of flooding.

5. REFRENCES

Alam, M., and Rabbani, M. (2007). *Vulnerabilities and responses to climate change for Dhaka*. Environment and Urbanization, 19(1), 81.

Bontempi. G., Taieb, S. B., and le Borgne, Y. A. (2012). *Machine learning Strategies for Time Series Forecasting*. In eBIBS, page 62-77. Springer

Clarke, R.T. (1973). *A Review of Some Mathematical Models Used in Hydrology, with Observations on their Calibration and Use*. Journal of Hydrology 19(1):1-20

Devia K.G., & Ganasri, B.P. and Dwarakish, G.S. (2015). *A Review on Hydrological Models*. Aquatic Procedia. 4. 1001-1007. 10.1016/j.aqpro.2015.02.126.

Douglas, I., Alam, K., Maghenda, M. A., Mcdonnell, Y., McLean, L., and Campbell, J. (2008). *Unjust waters: climate change, flooding and the urban poor in Africa*. Environment and Urbanization, 20(1), 187-205.

Eldho T. I. and Kulkarni A. T., (2017). *Conceptual and Physically Based Hydrological Modeling*. Book Chapter, Sustainable Water Resources Management

Jacobson C.R. (2011). *Identification and quantification of the hydrological impacts of imperviousness in urban catchments: A review.* J. of Environ. Management, 92(6): 1438-1448.

Ochoa-Rodriguez, S. et al. (2015). Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. Journal of Hydrology, 531, 389-407.

Parliamentary Office of Science and Technology (Parliamentary Office of Science and Technology) (2007), *Urban Flooding.* Postnote Number 289, Parliamentary Office of Science and Technology, London, (available at www.parliament.uk/documents/upload/postpn289.pdf).

Seyoum S.D. (2013). *Framework for Dynamic Modelling of Urban Floods at Different Topographical Resolutions.* PhD thesis, Delft University of Technology, Delft, The Netherlands.

Sudheer K. P., Nayak, P. C and Ramasastri K. S. (2003). *Improving Peak Flow Estimates in Artificial Neural Network River Flow Models*, Hydrological Process. 17, 677–686 (2003)

Verbeiren B., Seyoum S. D., Lubbad I., Xin T., Ten Veldhuis M.C., Onof C., Wang Li-Pen, Ochoa-Rodriguez S., Veeckman C., Boonen M., See L., Nalpas D., O'Brien B., Johnston A, and Willems P. (2018). *FloodCitiSense: Early Warning Service for Urban Pluvial Floods for and by Citizens and City Authorities*, 11th International Conference on Urban Drainage Modelling, 23-26, Sept, Palermo, Italy.

Wang, W. C., Chau, K. W., Cheng, C. T, and Qui, L. (2009). *A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series*. Journal of Hydrology. 374 (3-4):294-306